

(TM)

Distribution rights by Oxford Molecular Ltd

protein - protein database search, using Smith-Waterman algorithm

Thu Jan 18 09:51:49 2000; MasPar time 12.35 Seconds

generated.

US-09-332-063-2

416

MPRAQSSASYQVPADPFA.....KTPIQILGQEPDAEMVEYLI 675

9AM 150

## Chapter 11

30303 seqs, 12848679 residues

Minimum Match 08

isting first 45 summaries

-issued

1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

mean 34.900; Variance 263.923; scale 0.132

**pred.no.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	?	Score	Query Match	Length	DB	ID	Description	Pred. No
1	284	6.4	214	1	US-08-217-	Sequence 4, Applicatio	2.53e+0	
2	259	5.9	180	4	5273901-7	Patent No. 5273901.	4.99e+0	
3	259	5.9	180	4	5482709-6	Patent No. 5482709.	4.99e+0	
4	257	5.8	186	1	US-08-450-	Sequence 2, Applicatio	6.32e+0	
5	246	5.6	219	2	US-08-557-	Sequence 54, Applicati	2.32e+0	
6	245	5.5	234	1	US-08-642-	Sequence 51, Applicati	2.61e+0	
7	245	5.5	561	1	US-08-642-	Sequence 52, Applicati	2.61e+0	
8	241	5.5	907	3	PCT-US95-0	Sequence 19, Applicati	3.31e+0	
9	241	5.5	2441	1	US-08-194-	Sequence 2, Applicatio	4.19e+0	
10	238	5.4	262	2	US-08-929-	Sequence 1, Applicatio	5.96e+0	
11	238	5.4	262	1	US-08-403-	Sequence 1, Applicatio	5.96e+0	
12	238	5.4	263	2	US-08-557-	Sequence 51, Applicati	5.96e+0	
13	240	5.4	297	2	US-08-580-	Sequence 6, Applicatio	4.71e+0	
14	238	5.4	447	1	US-08-642-	Sequence 32, Applicati	9.54e+0	
15	234	5.3	330	1	US-08-642-	Sequence 65, Applicati	9.54e+0	
16	234	5.3	408	1	US-09-089-	Sequence 2, Applicatio	2.43e+0	
17	228	5.2	1248	2	US-07-812-	Sequence 2, Applicatio	1.93e+0	
18	226	5.1	98	2	US-07-812-	Sequence 2, Applicatio	1.93e+0	
19	226	5.1	98	2	US-07-814-	Sequence 2, Applicatio	1.93e+0	
20	227	5.1	777	1	US-08-642-	Sequence 53, Applicati	2.16e+0	
21	220	5.0	2035	2	US-08-479-	Sequence 2, Applicatio	4.90e+0	
22	215	4.9	475	2	US-08-396-	Sequence 14, Applicati	8.76e+0	
23	215	4.9	475	2	US-08-861-	Sequence 14, Applicati	8.76e+0	

45	195	4.4	144	1	US-08-642-	Sequence 49, Applicant	7.80e-04
44	197	4.5	723	4	5200183-4	Sequence 4, Applicant	7.80e-04
43	205	4.6	1255	2	US-09-080-	Sequence 2, Applicant	7.80e-04
42	209	4.7	2414	3	FCI-US95-9	Sequence 1, Applicant	7.80e-04
41	212	4.8	1867	2	US-08-227-	Sequence 2, Applicant	7.80e-04
40	214	4.8	1867	2	US-08-642-	Sequence 95, Applicant	7.80e-04
39	211	4.8	1077	1	US-08-419-	Sequence 5, Applicant	7.80e-04
38	211	4.8	1077	1	US-07-972-	Sequence 82, Applicant	7.80e-04
37	213	4.8	1026	2	US-08-614-	Sequence 7, Applicant	7.80e-04
36	217	4.9	1026	2	US-08-144-	Sequence 7, Applicant	7.80e-04
35	216	4.9	745	2	US-08-245-	Sequence 2, Applicant	7.80e-04
34	216	4.9	745	1	US-08-445-	Sequence 2, Applicant	7.80e-04
33	216	4.9	742	2	US-08-437-	Sequence 2, Applicant	7.80e-04
32	216	4.9	742	1	US-08-347-	Sequence 2, Applicant	7.80e-04
31	216	4.9	742	4	5200183-2	Patent No. 5200183	7.80e-04
30	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
29	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
28	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
27	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
26	216	4.9	722	1	US-08-347-	Sequence 1, Applicant	7.80e-04
25	216	4.9	722	4	5200183-3	Patent No. 5200183	7.80e-04
24	216	4.9	722	4	US-08-204-	Sequence 7, Applicant	7.80e-04
23	216	4.9	722	1	US-08-347-	Sequence 1, Applicant	7.80e-04
22	216	4.9	722	1	US-08-445-	Sequence 1, Applicant	7.80e-04
21	216	4.9	722	1	US-08-445-	Sequence 1, Applicant	7.80e-04
20	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
19	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
18	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
17	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
16	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
15	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
14	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
13	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
12	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
11	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
10	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
9	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
8	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
7	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
6	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
5	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
4	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
3	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
2	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04
1	216	4.9	722	1	US-08-445-	Sequence 3, Applicant	7.80e-04

## ALIGNMENTS

ID	US-08-217-327-4	STANDARD:	PRT;	214 AA.
XX				
XX				
XX	xxxxxx			
DE				
XX				
CC	Sequence 4, Application US/08217327			
CC				
CC	Patent No. 5474925			
CC	GENERAL INFORMATION:			
CC	APPLICANT: John, Makiyaka E			
CC	APPLICANT: Barton, Kenneth A			
CC	TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber			
CC	NUMBER OF SEQUENCES: 16			
CC	CORRESPONDENCE ADDRESS:			
CC	ADDRESSEE: Quarles and Brady			
CC	STREET: P.O. Box 2113			
CC	CITY: Madison			
CC	STATE: WI			
CC	COUNTRY: USA			
CC	ZIP: 53701-2113			
CC	COMPUTER READABLE FORM:			
CC	MEDIUM TYPE: Floppy disk			
CC	COMPUTER: IBM PC compatible			
CC	OPERATING SYSTEM: PC-DOS/MS-DOS			
CC	SOFTWARE: Patent In Release #1.0, Version #1.25			
CC	CURRENT APPLICATION DATA:			
CC	APPLICATION NUMBER: US/08/217,327			
CC	FILING DATE:			
CC	CLASSIFICATION: 435			
CC	PRIOR APPLICATION DATA:			
CC	APPLICATION NUMBER: US 07/812,233			
CC	FILING DATE: 19-DEC-1991			
CC	ATTORNEY/AGENT INFORMATION:			
CC	NAME: Seay, Nicholas J			
CC	REGISTRATION NUMBER: 27,386			
CC	REFERENCE/DOCKET NUMBER: 1122990831			
CC	TELECOMMUNICATION INFORMATION:			
CC	TELEPHONE: 608-251-5000			
CC	TELEFAX: 608-251-9166			
CC	INFORMATION FOR SEQ ID NO: 4:			
CC	SEQUENCE CHARACTERISTICS:			
CC	LENGTH: 214 amino acids			
CC	TYPE: amino acid			
CC	TOPOLOGY: linear			

CC SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CC CURRENT APPLICATION DATA:

Page 3

```
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MAKI, David J.
CC REGISTRATION NUMBER: 31,392
CC REFERENCE/DOCKET NUMBER: 210121.422
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (206) 632-4900
CC TELEFAX: (206) 682-6031
CC INFORMATION FOR SEQ ID NO: 54:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 219 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
SQ SEQUENCE 219 AA; 21347 MW; 161993 CN;

Query Match          5.6%; Score 246; DB 2; Length 219;
Best Local Similarity 37.9%; Pred. No. 2.32e-05;
Matches 58; Conservative 43; Mismatches 52; Indels 0; Gaps

Db      Dd      Qy      Db      Qy      Db      Qy
56 AAAAAAOKAAAKKAAAPSGSKSAAAAPAATAAAAKAAAPPARTAAAPAKAAAHAKA 115
   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
476 APVAAAAATTAATTATTITTTVAALPVNVAHAHAHAHAHAHAAAPSPATAAATAAVSPA 533
   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||

Dd      Db      Qy      Db      Qy
116 AAPPKAAAPAKPAKTAAAPKTAAPPAKAAAPPAKAAAPPAKAAAPPAKAAAPPAKAAAP 175
   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
536 GOIPAAASVSAAMAAVPSMAAAAVQVAPAAPAFVPAPALVPVAPAPAAQASAPAQGP 595
   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||

Db      Qy      Db      Qy
176 AKAAAPAKAAAPAKAAAPPAKAAAPPAEAP 208
   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
596 TSAAPVAPTPTPTPVAQAEEVPASPATGPG 628
   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 6 STANDARD; PRT: 234 AA.
ID US-08-642-255-51
AC xxxxxx
DT
XX
DE Sequence 51, Application US/08642255
XX
CC Sequence 51, Application US/08642255
CC Patent No. 5773249
CC GENERAL INFORMATION:
CC APPLICANT: CAPPELLO, Joseph
CC APPLICANT: FERRARI, Franco A.
CC TITLE OF INVENTION: High Molecular Weight Collagen-Like
CC TITLE OF INVENTION: Protein Polymers
CC NUMBER OF SEQUENCES: 135
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: EHLER, HOEBACH, TEST, ALBRITTON & HERBERT
CC STREET: 4 Embarcadero Center, Suite 3400
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94111-4187
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentia Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/642,255
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: ROWLAND, Berttram I.
CC REGISTRATION NUMBER: 20,015
CC REFERENCE/DOCKET NUMBER: AS5556-3/BIR
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 494-8700
CC TELEFAX: (415) 494-8771
CC
```



CC LENGTH: 907 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 907 AA; 94489 MW; 5007079 CN;

Query Match 5.5%; Score 243; DB 3; Length 907;  
Best Local Similarity 22.5%; Pred. No. 3.31e-05;  
Matches 50; Conservative 86; Mismatches 80; Indels 6; Gaps 6;

Db 471 ADVTSPTAGTSGASPTSPSPMDNGTE-SKAPDMSSTSPVTPPTPNATSPPTAVTT 529  
OY 429 PSTSPVSPSTPLSHSKSGSDCSTQTERGTEENKTAAPVAVPAAAAATAAAT 488  
Db 530 PTPNATSPPTAVTTPPTP-NATSPITGKTSPTSAVTPPTPNATSPPL-KTSPSTAVTPT 587  
OY 489 ATAATITTTTMAAAVAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 548  
Db 588 PNATSPITGKTSPTSAVTPPTPNAT-GPTVGETSPQANATNHTIGTSPPTVVTGPKN 645  
OY 549 AVAASAAAAAAVAPAPAPAPAPALVPAPAPAAQAASAPAQTAPTSAPAVAPTAPT 608  
Db 646 ATSAVTTGQHNITSSSTSMISRPSSNETLSPSTSDNSTH 687  
OY 609 PTPAVQAQEV-ASPATGPGPHRLSIPILCNPKTDGPFVH 649

RESULT 9  
ID US-08-194-468-2 STANDARD; PRT; 2441 AA.  
XX xxxxxx

DE Sequence 2, Application US/08194468  
CC Sequence 2, Application US/08194468  
CC Patent No. 5750336  
CC GENERAL INFORMATION:  
CC APPLICANT: Montmady, Marc R.  
CC TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF  
CC TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN  
CC NUMBER OF SEQUENCES: 3  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
CC STREET: 444 South Flower Street, Suite 2000  
CC CITY: Los Angeles  
CC STATE: California  
CC COUNTRY: USA  
CC ZIP: 90071  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/194,468  
CC FILING DATE: 10-FEB-1994  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Reiter, Stephen E.  
CC REGISTRATION NUMBER: 31,192  
CC REFERENCE/DOCKET NUMBER: P41 9672  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (619)-546-4737  
CC TELEFAX: (619)-546-9392  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 2441 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein

SQ SEQUENCE 2441 AA; 265473 MW; 30223014 CN;

Query Match 5.5%; Score 241; DB 1; Length 2441;  
Best Local Similarity 23.4%; Pred. No. 4.19e-05;  
Matches 58; Conservative 89; Mismatches 91; Indels 10; Gaps 10;

Db 764 SPSPPOPPNMGTHANNIMQAPQNFQPNQEPSSSGAMSVNMGCPAAGVSO 823  
OY 431 TPSEVPSTPLSHSKSGSDCSTQTERGTEENKTAAPVAVPAAAAATAAATA 489  
Db 824 GOERGAALPNLNLAPQASQLPCPEVYQSPPLHPTPPASTAGMPSLQHTAPGMTPTQ 883  
OY 490 TAAITTTTMAAAVAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 548  
Db 884 PAAPQSTPVSSQCTPTPTGVSVAQAQTSTPTVQAQAQVTPPOPPTVPSPVATP 943  
OY 549 AVADS-AAAA-AAVQVAPAPAPAPALVP-VPDA-PAAQAASAPQATAPAPAVA-P 603  
Db 944 QSSQOQPTPHVTPGPGPLGQAASIDNRVPTSTVSAETSSQCPGPDVPLEKTEVQ 1003  
OY 604 -TPAPTTPAVQAQ-EVPASPTGPGPHRLSIPS-LICNPDKTDGPFVHSLTERPTIQ 660  
Db 1004 TDDAEPP 1011  
OY 661 ILGQEPDA 668

RESULT 10  
ID US-08-929-414-1 STANDARD; PRT; 262 AA.  
XX xxxxxx

DE Sequence 1, Application US/08929414  
CC Sequence 1, Application US/08929414  
CC Patent No. 5942403  
CC GENERAL INFORMATION:  
CC APPLICANT: Reed, Steven G.  
CC APPLICANT: Houghton, Raymond  
CC APPLICANT: Skeiky, Yasir A.W.  
CC TITLE OF INVENTION: OF T. CRUZ1 INFECTION  
CC TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION  
CC NUMBER OF SEQUENCES: 15  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: SEED and BERRY LLP  
CC STREET: 6300 Columbia Center, 701 Fifth Avenue  
CC CITY: Seattle  
CC STATE: Washington  
CC COUNTRY: USA  
CC ZIP: 98104-7092  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/929,414  
CC FILING DATE: 15-SEP-1997  
CC CLASSIFICATION: 530  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Maki, David J.  
CC REGISTRATION NUMBER: 210121.406C1  
CC REFERENCE/DOCKET NUMBER: 210121.406C1  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (206) 622-4900  
CC TELEFAX: (206) 682-6031  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 262 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear





```

XX xx      xxxxxx
AC xx
DT xx
DE xx
XX xx
Sequence 32, Application US/08642255
CC CC      Sequence 32, Application US/08642255
CC CC      Patent No. 5773249
CC CC      GENERAL INFORMATION:
CC CC      APPLICANT: CAPPELLO, Joseph
CC CC      APPLICANT: FERRARI, Franco A.
CC CC      TITLE OF INVENTION: High Molecular Weight Collagen-Like
CC CC      TITLE OF INVENTION: Protein Polymers
CC CC      NUMBER OF SEQUENCES: 135
CC CC      CORRESPONDENCE ADDRESS:
CC CC      ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
CC CC      STREET: 4 Embarcadero Center, Suite 3400
CC CC      CITY: San Francisco
CC CC      STATE: California
CC CC      COUNTRY: USA
CC CC      ZIP: 94111-4187
CC CC      COMPUTER READEABLE FORM:
CC CC      MEDIUM TYPE: Floppy disk
CC CC      COMPUTER: IBM PC compatible
CC CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC CC      SOFTWARE: Patent Release #1.0, Version #1.30
CC CC      CURRENT APPLICATION DATA:
CC CC      APPLICATION NUMBER: US/08/642,255
CC CC      FILING DATE:
CC CC      CLASSIFICATION: 435
CC CC      ATTORNEY/AGENT INFORMATION:
CC CC      NAME: ROWLAND, Bertam I.
CC CC      REGISTRATION NUMBER: 20,015
CC CC      REFERENCE/DOCKET NUMBER: A55556-3/BIR
CC CC      TELECOMMUNICATION INFORMATION:
CC CC      TELEPHONE: (415) 494-8700
CC CC      TELEFAX: (415) 494-8771
CC CC      TELEX: 910 277299 FHT UR
CC CC      INFORMATION FOR SEQ ID NO: 32:
CC CC      SEQUENCE CHARACTERISTICS:
CC CC      LENGTH: 330 amino acids
CC CC      TYPE: amino acid
CC CC      STRANDEDNESS: single
CC CC      TOPOLOGY: linear
CC CC      MOLECULE TYPE: protein
CC CC      SEQUENCE 330 AA; 28876 MW; 625100 CN;
SO SO
ery Match          5.3%; Score 234; DB 1; Length 330;
at Local Similarity 22.1%; Pred. No. 9,54e-05;
Matches 49; Conservative 90; Mismatches 76; Indels 7; Gaps 7;
Db 36 PGPDPGPPGPGAPGCGPDPGPPGPPGAGPVGSFGABGPPGPPGPPGAPGPPG 95
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 429 PSTSPV-PPTSP-ILSHSKTSGSRDSTOTERGETESNKTAVAIVPAVAAATAA 486
   : : : | : | : | : | : | : | : | : | : | : | : | : | : | :
   : : : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 96 PGPPGPPGAPPVS-GCAPGPPGPPGPPGPPGAPGVSGABGPPGPPGPPGAPG 154
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 487 ITATRAATTITMVAALPYAVAAAAAPAAAAP-SPATAAATAAASVPAQAASA 545
   : : : | : | : | : | : | : | : | : | : | : | : | : | : | :
   : : : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 155 PGPDPGPPGADGPPGPPGPPGPA-GPVGSPGAPGPPGPPGPPGAPG-PPGP 212
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 546 SAAVAAPFAAAAVAVAPAAPVAPVALVPVAPAPAQAASNAQTGATSAPNAVPT 605
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 213 PGPDPGPPGAPVSGAPGPPGPPGPPGPPGAPGPPGPPG 254
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 606 APTTPAAQAEVPAASPATGP-GHRLISTSLTCNPDXTDGP 646

```

Search completed: Tue Jan 18 09:52:05 2000  
Job time : 16 secs.



\*\*\*\*\*  
MIPS RELEASE  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Bioinformatics Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_dp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jan 18 09:50:44 2000; MasPar time 43.38 Seconds  
849.350 Million cell updates/sec  
Juliar output not generated.

Title: >US-09-332-063-2  
Description: (1-675) from US09332063.pep  
Perfect Score: 4416  
Sequence: 1 MPRAOPSSASTQVPADPFA.....KTPIDILGCEPDAMKVEYLI 675

Scoring table:  
PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

## Database:

spiremle19  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle  
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 54.237; Variance 271.816; scale 0.200

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	372	8.4	507	13	013028 ANTIFREEZE GLYCOPOLYMER	1.18e-15
2	365	8.3	353	14	039266 COUNTERPART OF HSV-1 G	3.96e-15
3	359	8.1	807	5	018511 INSECT INTESTINAL MUCI	1.12e-14
4	341	7.7	867	14	039782 MEMBRANE GLYCOPROTEIN	2.47e-13
5	340	7.7	2378	5	091365 CODED FOR BY C. ELEGAN	2.93e-13
6	335	7.6	788	5	018510 INSECT INTESTINAL MUCI	6.88e-13
7	335	7.6	1439	1	039781 MEMBRANE GLYCOPROTEIN	6.88e-13
8	327	7.5	801	5	023635 SIMILAR TO LONG TANDEN	1.15e-12
9	327	7.5	216	6	028501 MUCIN (FRAGMENT)	2.69e-12
10	328	7.4	279	4	014888 MUCIN (FRAGMENT)	2.27e-12
11	321	7.3	1611	3	042854 HYPOTHEICAL 170.5 KD	7.45e-12
12	312	7.1	750	14	039307 POSITIONAL COUNTERPART	3.41e-11
13	307	7.0	770	5	020908 F56H9.1 PROTEIN.	7.92e-11
14	306	6.9	825	5	017921 C12D12.1 PROTEIN.	9.37e-11
15	304	6.9	831	5	017893 F53B11.3 PROTEIN.	1.31e-10
16	300	6.8	473	10	039620 (VSP-3) PRECURSOR.	2.56e-10
17	302	6.8	1188	10	041805 EXTENSIN-LIKE PROTEIN	1.83e-10
18	297	6.7	729	2	069743 HYPOTHEICAL 74.5 KD P	4.23e-10
19	290	6.6	581	5	020517 F47B8.5 PROTEIN.	1.36e-09
20	292	6.6	3247	14	065553 UT56.	9.75e-10

RESULT	ID	Score	Query Match	Length	DB ID	Description	Pred. No.
21	288	6.5	167	4	014883	INTESTINAL MUCIN (FRAG	1.90e-09
22	288	6.5	503	14	039779	SEGMENT PROTEIN (FRAG	1.90e-09
23	285	6.5	528	6	029071	GASTRIC MUCIN (FRAGMEN	3.12e-09
24	287	6.5	600	5	021036	CODDED FOR BY C. ELEGAN	2.24e-09
25	284	6.4	214	10	039789	PROLINE-RICH CELL WALL	3.68e-09
26	283	6.4	214	10	039763	PROLINE-RICH CELL WALL	4.34e-09
27	282	6.4	602	6	019115	MUCI (FRAGMENT)	5.13e-09
28	278	6.3	169	10	041071	ARABINOGLACTAN-LIKE P	9.92e-09
29	280	6.3	197	10	082327	POTATIVE PROLINE-RICH	7.13e-09
30	278	6.3	317	6	029070	GASTRIC MUCIN (FRAGMEN	9.92e-09
31	279	6.3	483	14	083344	GLYCOPROTEIN 150.	8.42e-09
32	280	6.3	722	13	013083	ARGP POLYPROTEIN PRECU	7.13e-09
33	278	6.3	734	4	075112	KIA0613 PROTEIN (FRAG	9.92e-09
34	268	6.3	278	4	014884	INTESTINAL MUCIN (FRAG	5.13e-09
35	270	6.1	351	10	039492	W6 PRECURSOR.	3.70e-08
36	271	6.1	447	11	063348	MUCIN (FRAGMENT).	4.36e-08
37	269	6.1	464	10	041645	EXTENSIN (FRAGMENT).	4.36e-08
38	270	6.1	1151	13	057580	HIGH MOLECULAR MASS NU	3.70e-08
39	266	6.0	228	10	043558	PROLINE RICH PROTEIN P	7.12e-08
40	264	6.0	362	13	090400	ARGP POLYPROTEIN PRECU	9.87e-08
41	263	6.0	901	2	044562	STALDASE.	1.16e-07
42	259	5.9	236	4	014761	SMALL INTESTINAL MUCIN	2.23e-07
43	262	5.9	322	11	062605	APOMUCIN PRECURSOR.	1.37e-07
44	261	5.9	1404	4	092954	MEGAKARYOCYTE STIMULAT	1.61e-07
45	259	5.9	1777	5	022579	SIMILAR TO A. FAECALIS	2.23e-07

## ALIGNMENTS

RESULT	ID	Score	Query Match	Length	DB ID	Description	Pred. No.
1	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
2	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
3	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
4	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
5	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
6	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
7	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
8	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
9	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
10	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
11	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
12	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
13	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
14	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
15	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
16	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
17	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
18	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
19	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
20	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
21	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
22	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
23	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
24	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
25	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
26	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
27	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
28	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
29	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
30	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
31	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
32	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
33	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
34	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
35	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
36	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
37	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
38	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
39	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
40	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
41	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
42	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
43	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
44	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	
45	013028	8.48	507	AA: 44684	MM: 675D6FCB	CRC33;	



[illegible]

```

RESULT 9
028501 PRELIMINARY: PRF: 216 AA.
028501;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MUCIN (FRAGMENT).
GN MUC2.
OC MACACA MULATTA (RHESUS MACAQUE).
OC EUKARIOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CARNIVORINI; CERCOPIITHECIDAE; CERCOPIITHECINAE; MACACA.
RN [1]
RN SEQUENCE FROM N.A.
RP
RC TISSUE=TRACHEA;
RX MEDLINE: 94235322.
RA AN G., LUO G., WU R.;
RT "expression of MUC2 gene is down-regulated by vitamin A at the
transcriptional level in vitro in tracheobronchial epithelial
cells.";
RL AM. J. RESPIR. CELL MOL. BIOL. 10:546-551(1994).
DR ENBL: U00483; G437055; .
FT NON_TER 1
SQ SEQUENCE 216 AA: 21692 MW; 906FCB59 CRC32;

Query Match 7.4%; Score 327; DB 6; Length 216;
Best Local Similarity 22.0%; Pred No. 2,69e-12;

```

```

Db      54  TTPRPV-TVSTSSPTSTTSAPPTSTTSASTSTSGGCTPPSPVPTSTTSAPTPT 112
      :|| ||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
Qy      430 STSPSVPRSTLHAHKTGRCDSCTEGTENGTSNKTAAPAPTSVPAPVAAAAITAAITA 489
Db      113 TSAPISST-TSATTTSTTSGCGTTPSPVPTSTTSAPTSTT-SGPTTPSAVPTTS 170
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Qy      490 TTAATTMTMAAAVAVAAAAAPAAAAAPSPATAATAAAAVSPRAAQOTPAALASVASMAA 549
Db      171 VPTTSTNSAPISSTTSATTTSRISGPTTSPVPTASTSTTSAGTTPSPVPT- 229
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Qy      550 VAPSAAAAAAQVAPAAPAPPPAPALPVVAPAPAAQAOTAPLSAPAVAPTPAPTP 609
Db      230 TSTISVPTSTTSASTSTTSASPTSTTSGTTPSPV 267
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Qy      610 TPVAAQAEVAPSPATGCPHNLSTPSLTCHNDKTCGEV 647

```

OC	RESULT	PR	AA
00	11		
01	042854	PRELIMINARY	1611
02	042854		
03	01-JUN-1998	(PREMREL	06
04	01-JUN-1998	(PREMREL	06
05	01-JUN-1998	(PREMREL	06
06	01-JUN-1998	(PREMREL	06
07	01-JUN-1998	(PREMREL	06
08	01-JUN-1998	(PREMREL	06
09	01-JUN-1998	(PREMREL	06
10	01-JUN-1998	(PREMREL	06
11	01-JUN-1998	(PREMREL	06
12	01-JUN-1998	(PREMREL	06
13	01-JUN-1998	(PREMREL	06
14	01-JUN-1998	(PREMREL	06
15	01-JUN-1998	(PREMREL	06
16	01-JUN-1998	(PREMREL	06
17	01-JUN-1998	(PREMREL	06
18	01-JUN-1998	(PREMREL	06
19	01-JUN-1998	(PREMREL	06
20	01-JUN-1998	(PREMREL	06
21	01-JUN-1998	(PREMREL	06
22	01-JUN-1998	(PREMREL	06
23	01-JUN-1998	(PREMREL	06
24	01-JUN-1998	(PREMREL	06
25	01-JUN-1998	(PREMREL	06
26	01-JUN-1998	(PREMREL	06
27	01-JUN-1998	(PREMREL	06
28	01-JUN-1998	(PREMREL	06
29	01-JUN-1998	(PREMREL	06
30	01-JUN-1998	(PREMREL	06
31	01-JUN-1998	(PREMREL	06
32	01-JUN-1998	(PREMREL	06
33	01-JUN-1998	(PREMREL	06
34	01-JUN-1998	(PREMREL	06
35	01-JUN-1998	(PREMREL	06
36	01-JUN-1998	(PREMREL	06
37	01-JUN-1998	(PREMREL	06
38	01-JUN-1998	(PREMREL	06
39	01-JUN-1998	(PREMREL	06
40	01-JUN-1998	(PREMREL	06
41	01-JUN-1998	(PREMREL	06
42	01-JUN-1998	(PREMREL	06
43	01-JUN-1998	(PREMREL	06
44	01-JUN-1998	(PREMREL	06
45	01-JUN-1998	(PREMREL	06
46	01-JUN-1998	(PREMREL	06
47	01-JUN-1998	(PREMREL	06
48	01-JUN-1998	(PREMREL	06
49	01-JUN-1998	(PREMREL	06
50	01-JUN-1998	(PREMREL	06
51	01-JUN-1998	(PREMREL	06
52	01-JUN-1998	(PREMREL	06
53	01-JUN-1998	(PREMREL	06
54	01-JUN-1998	(PREMREL	06
55	01-JUN-1998	(PREMREL	06
56	01-JUN-1998	(PREMREL	06
57	01-JUN-1998	(PREMREL	06
58	01-JUN-1998	(PREMREL	06
59	01-JUN-1998	(PREMREL	06
60	01-JUN-1998	(PREMREL	06
61	01-JUN-1998	(PREMREL	06
62	01-JUN-1998	(PREMREL	06
63	01-JUN-1998	(PREMREL	06
64	01-JUN-1998	(PREMREL	06
65	01-JUN-1998	(PREMREL	06
66	01-JUN-1998	(PREMREL	06
67	01-JUN-1998	(PREMREL	06
68	01-JUN-1998	(PREMREL	06
69	01-JUN-1998	(PREMREL	06
70	01-JUN-1998	(PREMREL	06
71	01-JUN-1998	(PREMREL	06
72	01-JUN-1998	(PREMREL	06
73	01-JUN-1998	(PREMREL	06
74	01-JUN-1998	(PREMREL	06
75	01-JUN-1998	(PREMREL	06
76	01-JUN-1998	(PREMREL	06
77	01-JUN-1998	(PREMREL	06
78	01-JUN-1998	(PREMREL	06
79	01-JUN-1998	(PREMREL	06
80	01-JUN-1998	(PREMREL	06
81	01-JUN-1998	(PREMREL	06
82	01-JUN-1998	(PREMREL	06
83	01-JUN-1998	(PREMREL	06
84	01-JUN-1998	(PREMREL	06
85	01-JUN-1998	(PREMREL	06
86	01-JUN-1998		

```

ID RESULT 12
AC 039307 PRELIMINARY: PRT; 750 AA.
DT 039307:
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE POSITIONAL COUNTERPART OF HSV-1 GENE US5.
DE
DE 71.
OS EQUINE HERPESVIRUS 4.
OC VIRUSES; DSDNA VIRUSES; NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.
RN [1]
RN SEQUENCE OF 685-750 FROM N.A.
RN STRAIN-NS80567;
RN MEDLINE: 93389454.
RN CULLINANNE A.A., NEILAN J., WILSON L., DAVIDSON A.J., ALLEN G.;
RN "The DNA sequence of the equine herpesvirus 4 gene encoding
RN glycoprotein gp17/18, the homologue of herpes simplex virus
RN glycoprotein gD.";
RN J. GEN. VIROL. 74:0-0(0).
RN [2]
RN SEQUENCE OF 1-111 FROM N.A.
RN STRAIN-NS80567;
RN MEDLINE: 93119267.
RN NAGESHA H.S., CRABB B.S., STUDDERT M.J.;
RN "Analysis of the nucleotide sequence of five genes at the left end of
RN the unique short region of the equine herpesvirus 4 genome.";
RN ARCH. VIROL. 128:143-154(1993).
RN [3]
RN SEQUENCE FROM N.A.
RN STRAIN-NS80567;
RN MEDLINE: 98264497.
RN TELFORD E.A.R., WATSON M.S., PERRY J., CULLINANNE A.A., DAVIDSON A.J.;
RN "The DNA sequence of equine herpesvirus 4.";
RN J. GEN. VIROL. 79:1197-1203(1998).
RN [4]
RN SEQUENCE FROM N.A.

```

RP SEQUENCE: FROM N.A.  
RX MEDLINE: 94150718.  
RA WILSON R., AINSOUGH R., ANDERSON S., BAYNES C., BEERS M.,  
RA BONFIELD J., BIRTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,  
RA CRAWTON M., DEAR S., DU Z., DUREIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HARKINS T., HILIER L., JER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIKKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VADDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOLDMAN P.,  
RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*  
RT *elegans*."   
RL NATURE 368:32-38(1994).  
DR EMBL: Z74473; E1347264; -  
SQ SEQUENCE 770 AA; 81625 MW; 771AB928 CRC32;

[illegible]

RESULT	15		
ID	017893	PRELIMINARY;	PRT; 851 AA

Search completed: Tue Jan 18 09:51:31 2000  
Job time : 47 secs.

...DEARN...SVKG--LIQVEDGK 2610

RESULT 4  
US-08-403-379A-1  
; Sequence 1, Application US/08403379A  
; Patent No. 5756662  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION  
; TITLE OF INVENTION: OF T. CRUZI INFECTION  
; NUMBER OF SEQUENCES: 9

32-06:

CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,379A  
FILING DATE: 14-MAR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.406  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 262 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-403-379A-1

Query Match 6.4%; Score 213; DB 1; Length 262;  
Best Local Similarity 48.4%; Pred. No. 2.8e-07;  
Matches 74; Conservative 9; Mismatches 54; Indels 16; Gaps 9;

QY 479 AAAATAAAITATAATITTTMVAAP---VAVAAAAAPA-AAAAPSPATAAAATAAVSPAA 534  
Db 112 AAAAATAAKQKAAAK-----AAAPSGKKSAAAIAPAKAAAAAPAKAAAAAPAK 165  
QY 535 AGQIPAAASVASA-AAVAPSAAAAAVQVAPAPVPAPALVPVPAPAAQAQAPAQ 593  
Db 166 AAAAPAKAAAAAPAKAATAPAKAAAAAPAKTA-AAAPAKAAAPAKA-AAAPAKA-ATAPAKAA 222  
QY 594 AFTSAPAVAPTPTPTTPAVAQAQEVPAATGP 626  
Db 223 AAPAKAATAPAKAAT-APAKA-AAAPAKAATAP 253

RESULT 5